Extending the Linking Capabilities of Resources in the Web of Data

Bujar Raufi, Xhemal Zenuni, and Artan Luma,

***Abstract*—This paper investigates the possibility of extending web resource linkings with other linking methodologies besides the unidirectional one. The process of interlinking web resources nowadays is done through hypertext references which allow only unidirectional coupling. We explore the extension of the linking process in the web of data through the use of XML Linking Technologies (XLink) applied to web documents or other resources. With the proposed approach, the linking structure of web resources can be substantially enriched and as a result, access to more resources and resource entities is achieved.**

The initial experimental results also indicate that by using the proposed approach, a substantial improvement and increase of linking process between web resources is evident. We envision that this outline will potentially improve future of information extraction and resource linking for automatically generate knowledge bases and various assertions on such linked resources.

***Keywords*—WWW, Web of Data, Linked Open Data, XLink, Semantic Web**

I. Introduction

THe process of linking in WWW today is based solely on unidirectional links generated through hypertext's `<a>...</a>` (anchor) tag. This has represented the simplest and easiest way of inserting links to refer in either documents, images or a particular document section in hypertext. However, this way of unidirectional linking does not allow much flexibility in the sense of resource navigation, search and content linking. The main drawbacks of unidirectional links can be summarized as follows:

- The process of linking and referencing is solely in one direction (unidirectional). This means that the browsing is done solely on the feed-forward fashion.
- It refers to a single document (resource) with a single click even though that related documents can be available around the resource and can be displayed to users.
- The search of such resources is limited due to inability of linking more resources and resource sections at once.
- It severely impacts the fourth principle of Linked Open Data [1] which states that in order to discover more resources; more links should be included so that additional resources and things can be discovered.

In order to address the above mentioned issues, an extension of the linking process for the web in general and in the web of data in particular, is presented with the adoption of XML Linking Technologies (XLink) [2] applied to web resources. With this process, the link structure of web resources can be substantially enriched and as a result, the access to more resources is achieved. This approach makes the data on the web machine-processable as well as human-readable.

The process of link enrichment is completed with extension of simple links already used in hypertext resources with other types of links offered through XLink, which together with proprietary XML technologies such as XSL(T) [3] which transforms such links and renders these resources human-readable as well as user friendly.

The rest of the paper is organized as follows: section II focuses around related works done up to now related to the enrichment of links; section III elaborates the basic ideas of XLink and the Web of Data; section IV brings to an issue our methodology used for enrichment of the link content structure for the web of data together with some intitial experimental results that tests the groundess of our approach and section V concludes this paper together with some future works and directions.

II. Related Work

In relation to resource and entity linkings, many systems have been developed in consideration to bridge the gap between other formats and XLink, however, to the best of our knowledge, no particular effort has been given to the use of underlying technologies that are widely supported by browsers in furtherance of enriching the actual link structure in web resources with XLink.

One of many systems developed to use XLink for enriching and augmenting the link structure on the web is XSpect which represent an XLink based hypermedia system. The system is comprised of a set of XSLT stylesheets characterizing the transformation from Open Hypermedia Interchange Formats - OHIF files to XLink linkbases and vice versa. The system is also equipped with XSL Stylesheets that transform the XLink linkbases to an HTML representation that may be delivered to web browsers [4]. The problem with XSpect is that it offers more general opportunity to annotate and link document sections with more extended link types that XLink strives, yet it is still a platform, client and server dependent. In order to test and use XSpect, either the client version should be downloaded or the server version should be visited. Another drawback of XSpect is the predefined XLinks as a result of the transformation from Open Hypermedia Interchange Formats - OHIF files to XLink linkbases or HTML and vice versa.

An alternate tool that has been developed primarily for web editing is Amaya [5]. Amaya represent a tool which is

B. Raufi is with the Department of Computer Science - Faculty of Contemporary Sciences and Technologies , South East European University, Tetovo, 1200, Macedonia e-mail: (see http://www.seeu.edu.mk/ b.raufi).

Xh. Zenuni and A. Luma are with the Department of Computer Science - Faculty of Contemporary Sciences and Technologies , South East European University, Tetovo, 1200, Macedonia e-mail:(xh.zenuni,a.luma@seeu.edu.mk)

Manuscript received November 10, 2013; revised January 11, 2014.

utilized to create, update and annotate documents directly on the Web. Browsing capabilities are integrated with the editing and remote access features that are possible in an integrated fashion. Amaya was developed as joint effort between INRIA [16] and W3C [17] with the purpose of facilitating usage of new technologies for generating valid web pages. The main problem that arises in Amaya is that it does not offer direct support for XLink in order to create more sophisticated links that XLink support such as extended links (locators and arcs) as well as linkbases. Another system that tends to bridge the gap between XLink usage and its adoption to Semantic Web is XSRS [13]. The main idea inside SXRS is developing Semantic XLink Recommendation System (SXRS) which makes use of underlying semantic web technologies and is composed of XLink base, Knowledge base, Search Engine, and Inference Engine to provide three different approaches to represent the linking knowledge.

The main drawbacks of all of the above mentioned systems can be summarized in three fundamental aspects. The first aspect is that they lack platform independency, where some of the systems require the client or server version of their application which should be downloaded and used. The second aspect is the lack of the usage of general and standardized formats that makes it widespread adoption hard even though such direct format like XLink exists. In XSRS on the other hand, there is no clear method which shows in what way a connection between low level data and semantic level is provided.

Another problem that relates to resource and entity linkings at large scale is that interlinking of thousands and even millions of resources result in systematic interlinking errors in cases when some of the resources are linked many times than expected [14]. The adption of XLink technologies would substantially bridge this gap by the adption of *Extended links* and *Linkbases*.

Finally, some of the systems fail to offer extended link support which substantially contributes to link enrichment and augmentation. The aim of this paper is that through direct use of XLink technology to transform the documents or resources with XSL(T) to hypertext in order to achieve link enriching in the web. It is worth mentioning that in the recent years many browsers are adopting XSLT transformation as built-in capabilities which in the future would substantially nurture platform independency for the technology.

III. Linking Structure on the Web of Data

Linked Data or the Web of Data is concentrated around using the Web to connect related data that was not previously linked, or using the Web to lower the barriers on linking data that is currently linked using other methods. This would make the exposed data machine processable as well as human-readable. More specifically, Linked Open Data can be defined as a recommendation of best practices for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and proprietary data formats (RDF, N3, Turtle etc..). In recent years we are witnessing a tremendous movement from the document-centric concepts and technologies towards data-centric orientation where the Web of Documents tend to be replaced with the Web of Data. Populating the web with raw data and RDF links among them is the main idea behind the Linked Open Data or Web of Data principle. The Web of Data integrates various knowledge and data starting from Wikipedia, governmental and geospatial data, bioinformatics, science, publications and research up until entertainment. The idea of Web of Data gains a serious momentum in the scientific community; this is illustrated by the fact that in October 2007 when the project was initiated by Tim Berners-Lee, datasets consisted of over 2 billion RDF triples, which were interlinked by over 2 million RDF links. By September 2010 this had grown to 25 billion RDF triples, interlinked by around 395 million RDF links and by September 2011 this had grown to 31 billion RDF triples, interlinked by around 504 million RDF links [1],[6]. Its growth can be clearly seen from the linked open data cloud [10] as depicted in figure 1 The basic principle of the Web of Data lies on the

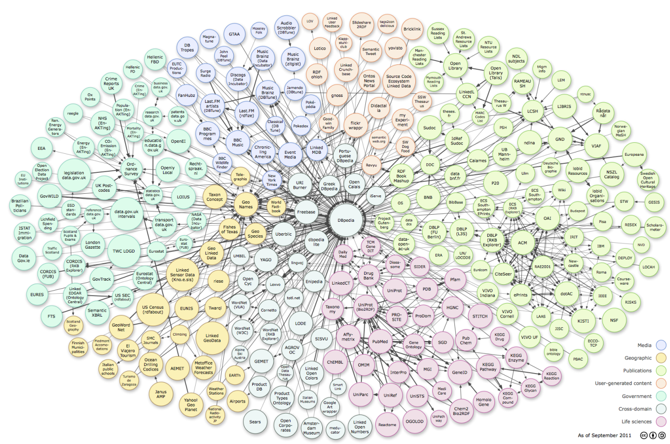

Fig. 1. Link Open data Cloud as of September 2011 (Cyganiak & Jentzsch)

underlying semantic web principles and is not concentrated solely on importing and generating data on the web, but it is also about generating large number of links through which users can explore more data. Therefore, when exploring information space on the web we find the necessary relevant data which also might link to other relevant resources. What is characteristic for the Web of Data is that like the ordinary web it is constructed of resources. The difference rests in the ability of the Web of Data to create links not only between documents, but also among things and other fine grained concepts of data. While the ordinary web uses hypertext documents through (X)HTML to create links with the anchor (`<a>...</a>` tag), the web of data uses the same anchors with URI's to link resources described by RDF or some other similar formats. The URIs tends to identify any kind of object or concept. Berners-Lee recommended the four rules principle required for rendering the web of data expansive such as[8]:

1) Using URIs as names for things
2) Utilizing HTTP URIs so that people can look up those names.
3) By looking up URI's, useful information should be

provided, using the standards (RDF, SPARQL)

4) Including links to other URI's so that user can discover more things

The web of data is based on two basic mechanisms of resource interlinking (both done through RDF) [15]:

1) through `owl:sameAs` construct which creates identity linkings between various RDF resources
2) with the use of domain specific constructs such as `foaf:knows`, `dc:author`, `dc:title` etc

It is worth mentioning that both approaches provide unidirectional linkings and offer no support for extended and linkbase types available in XLink.

Based on the above mentioned constraints of unidirecionality of the web of data as well as the request that access to more resources should be provied, we propose a specific methodology based on XML Linking Language (XLink) and XML Transformation Languages (XSLT) for transforming the XLink resources into viable hypertext links which in itself guard the semantic information for each link given in XLink files. Their access is still done through traditional links used in web due to the constraining nature of the same in this sense. However, from the machine processing point of view, each RDF resource in the web of data is enriched with other type of links that XLink can assign as it will be elaborated in the following section.

IV. Enriching Link Structure with XLink

One of the technologies that ensure extension of simple linking used in ordinary hypertext is XLink [2]. Xlink represent an attribute based syntax used for linking resources. Despite the fact that XLink can be simple relationships between two points A and B, they can also represent bidirectional relationships between resources, i.e from point B to A as well or in some cases multidirectional relationships are possible by linking many resources from a single point of origin altogether at the same time.

A. Simple and Extended Links in XLink

A simple link is characterized as a one-way connection between two resources. The starting source of the connection itself can be an XML document or a particular resource which further on can be transformed into any other format such as (X)HTML and the target or ending resource of the connection is identified by a Uniform Resource Identifier (URI). The link follows the route from the starting resource to a designated ending resource where a starting resource is an XML element that can be transformed into any other format and the ending resource may be an XML document, a particular section of an XML document, a group of elements in an XML document, a span of text in an XML document, RDF resource, or some other format that is not directly related to XML, such as an MPEG movie, a PDF file, CSV, Geo Location file etc. Below is illustrated a simple link example that can be can be added for a particular resource or object in the web of data: In the overhead illustrated XML, three attributes are important when it comes to XLINK: the namespace (`xmlns:xlink`), the type of link (`xlink:type`) and the reference to the resource (`xlink:href`).

Listing 1: A Simple Link with XLink

```
<object xmlns:xlink= "http://www.w3.org/1999/xlink"
        xlink:type = "simple"
        xlink:href = "http://www.imdb.com/title/tt1911644/">
  <type>Movie</type>
  <title>The Call</title>
  <category>Thriller</category>
  <director>Brad Anderson</director>
  <year>2013</year>
</object>
```

Whilst a simple links relate one resource (XML document) to another resource in a unidirectional fashion, an extended link describes connections between a set of resources and collection paths between those resources where each path connects exactly two resources. Any individual resource may be connected to one, two, none or all of the other resources, or any subset of the other resources available in the collection. A collection may be an intranet repository or even a Linked Open Data Cloud. A resource may even be connected back to itself. From a computer science perspective, extended links represent a directed, labelled graph in which the paths are arcs, the resources are vertices, and the URI's are labels. Extended links are of type *"locator"* which identify various resources for the same XML data identifiable by an URI and of an *"arc"* type to identify the order of paths between such resources. An example of extended links given for the above mentioned movie database regarding the review of the movie would look like as in listing IV-A. Another interesting extended links that preserves the rules for traversing among its participating resources by means of a set of elements are the *arcs*. The arc elements as part of extended links define the rules in what way every particular resource is traversed as depicted in 2. This diagram reflects a partially directional traversal arcs created by the facts that both Y and Z are allowed to initiate traversal to all X and partially to Y through Z. Because some labels appear on several resources, each arc specification potentially creates several traversal arcs at once. Arc elements are created

Listing 2: Extended Links Written in XLink

```
<object xmlns:xlink= "http://www.w3.org/1999/xlink"
        xlink:type = "extended">
  <type>Movie</type>
  <title>The Call</title>
  <category>Thriller</category>
  <director>Brad Anderson</director>
  <year>2013</year>
  <link xlink:type="locator"
        xlink:href="www.rottentomatoes.com/m/the_call_2013/"
        xlink:title="Rotten Tomatoes"
        xlink:role="http://purl.org/dc/elements/1.1/publisher"
        xlink:label="Reviewer"/>
 </link>
 <link xlink:type="locator"
       xlink:href="http://www.filmratings.com/
                  search.html?filmTitle=the+call&x=-1106&y=-41"
       xlink:title="Film Ratings"
       xlink:role="http://purl.org/dc/elements/1.1/publisher"
       xlink:label="Reviewer"/>
</link>
</object>
```

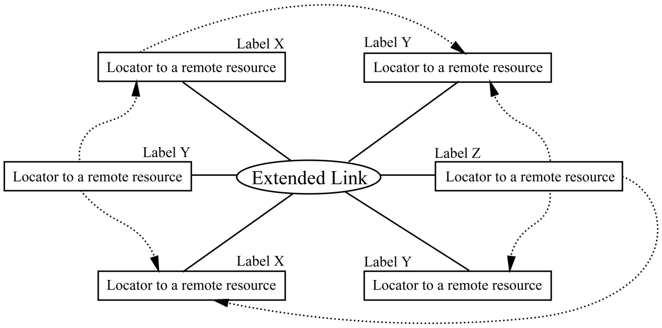

Fig. 2. Extended links with arc elements

by definition of `arc` value in `type` attribute and by previously defining labels for each extended link.

Current web browsers at most support simple XLinks that do little more than duplicate the functionality of hypertext's `<a></a>` element and almost no support for multidirectional or extended links [6]. Many browsers, including Internet Explorer, don't support XLinks at all and consequently, there is no mechanism for interpreting them as links other than simple anchor tags. The idea in this paper is to use the underlying web technologies and browser support to emulate the extended links. This is done through the mechanism of XSLT transformations of XML documents together with AJAX and JQuery [9] scripting technologies. The process of this transformation and enrichment of link structure in the web of data is depicted as in figure 3. From the figure, it can

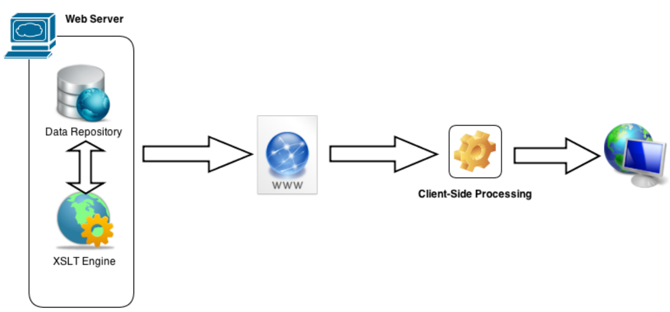

Fig. 3. Transformation and enrichment of link structure

be seen that the process link enrichment starts from data repository which can be consisted of any RDBMS models or simple XML documents in which the data and their respective XLink metadata resides. The XSLT engine transforms the XML document in its respective format and AJAX which represents the links in a human readable format. Listing IV-A illustrates a glimpse of the code which transforms the data repository of particular movie database objects together with their respective editions into a web processable hypertext in (X)HTML format. As a consequence, the user experience of link enrichment is illustrated as in figure 4. From the above figure, it is clearely observable that from a single link that user clicked, several options for that particular object or resource emerge, which on the semantic level simulate the extended link

Listing 3: Transformation of data Repository into (X)HTML with XLink metadata

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:xlink = "http://www.w3.org/1999/xlink">
  <xsl:template match="/"><xsl:apply-templates/></xsl:template>
  <xsl:template match="objects">
<xsl:apply-templates select="object"/>
  </xsl:template>
  <xsl:template match="object">
    <span class="item">
<span class="style">Movie name:</span>
    <span class="object"><xsl:value-of select="name"/>
    </span><br/>
 Director: <xsl:value-of select="director"/><br/>
 Release date:<xsl:value-of select="released"/><br/>
 Reviews:<br/><xsl:apply-templates select="link"/><br/>
</span>
</xsl:template>
<xsl:template match="link">
 <a xlink:show="new" xlink:type="locator" xlink:href="{@href}"
    onclick="location.href='{@href}'">
    <xsl:value-of select="title"/></a><br/>
</xsl:template>
</xsl:stylesheet>
```

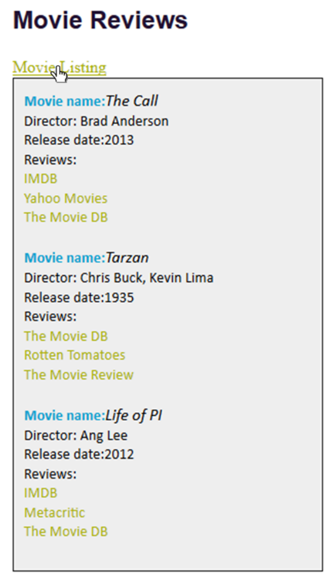

Fig. 4. User experience from link enrichment

category available in XLink. The advantage of this approach lies in the fact that no prior plugins or any additional tools need to be installed all that user needs is a browser that supports JavaScript and JQuery.

B. Experimental Results

Considering that our system is suitable for both documents stored in some particular formats (XML or (X)HTML) or some specific format notation for resource description (such as RDF), some experimental results regarding the percentage of link enrichment can be derived. If we consider a webmaster's guideline that suggests that the number of links in a single page should not exceed 100 [12], and if we take into consideration that for a particular link minimum of 3 linkbases can be attached, than experimental results indicate that in extended links we have a minimum of 34% enrichment seen from a page perspective as a whole.

Table I shows the comparison of various types of XLink enrichment for both ordinary web and the web of data

TABLE I. XLINK ENRICHMENT FOR BOTH WWW AND WEB OF DATA

	Type	Structure	Regular links	XLink
WWW	Documents	Semi-structured	<100	294
Web of Data	Resources	Structured	aprox. 540M	1.58B

Figure 5 shows the correspondence and the link enrichment via XLink in comparison to ordinary unidirectional links used in hypertexts.

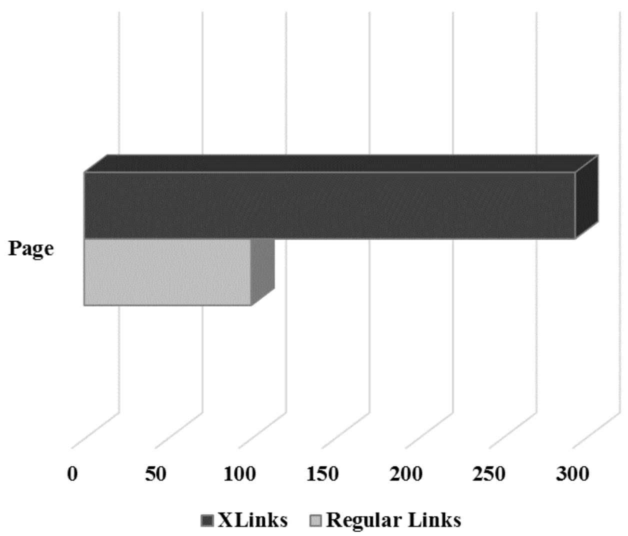

Fig. 5. Link Enrichment with XLink

Another testbed from where tangible results can be obtained for our approach lies in the work of (Mausam&Etzioni) [13] where context matching by using more sources increase the entity linking accuracy around 50.64%. If we adopt this approach and we take into consideration as above that for a particular link minimum of 3 linkbases can be attached, than experimental results asserted from the work of (Mausam&Etzioni) gives some aditional 34.5 to the context matching increase when using XLink.

Table II depicts the link accuracy between ordinary Web and the web of data when using XLink for link enrichment. It is obvious that from the overall link accuracy expectations we have substracted the average unlikable antities that always apear during linking process which was around 33% [14]. Figure 6 illustrates the link accuracy of XLink utlilized in the web of data in comparison to standard unidirectional web.

We foresee that the indicated experimental results give e promissing starting point that potentially could improove future of information extraction and resource linkings for automatically generating knowledge bases and various assertions on such linked resources in the web of data. However, in order to conclusively support it, additional experimental results and testings should be conducted.

TABLE II. LINK ACCURACY COMPARED TO WWW [14] AND LOD

String Link Ambiguity	**¡0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**	**0.9-1**
Link Accuracy (WWW)	2.37	3.31	8.54	12.95	23.47
Excpected link Accuracy(LOD)	7.11	9.93	25.62	38.85	70.41
Unlinkable Entities	0.23463	0.32769	0.84546	1.28205	2.32353
Relative Gain to Link Accuracy (LOD)	6.87537	9.60231	24.77454	37.56795	68.08647

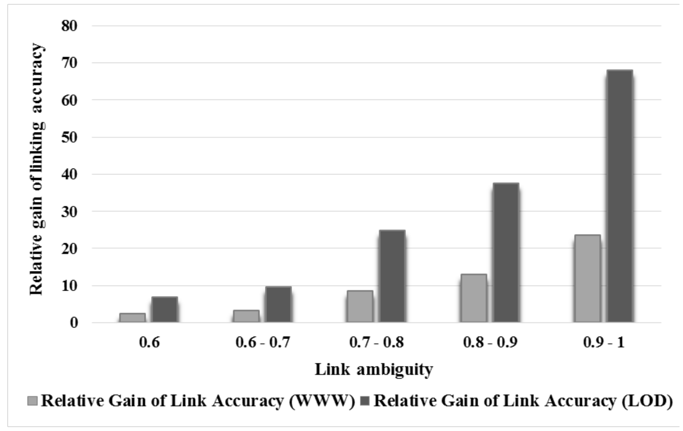

Fig. 6. Link Enrichment with XLink for LOD in comparison to WWW especially in cases where link ambiguity tend to be high

V. CONCLUSION AND FUTURE WORK

In this paper we have illustrated a methodology of link enrichment in the web of data where different resources can be improved with additional hyperlinks which was not possible in classical hypertext cases which supports only simple (unidirectional) links. Furthermore this approach does not obstruct the user experience while browsing and additionally embellishes the user navigation paths with more links. We have also demonstrated the implications of XLink usage in the enrichment process for both ordinary web and the Web of Data.

Future work regarding the link enrichment would involve:

1) Creating cross-browser support for XSL(T) transformations of XML files for link enrichment. For the time being, only Mozilla Firefox has a good support for XSL(T).
2) Allowing not only XML documents, but other data formats such as RDF or OWL Vocabularies to be able to be linked through XLink and rendered user-friendly through XSL(T) and Ajax/JQuery. This would be a major breakthrough, considering that all the web of data is based on URI's and resource descriptions through RDF.

3) Exploring the possibilities of visualisation extended link connections (link bases, resources and arcs in XLink) between various resources in the web of data in order to make the browsing process more user-friendly.
4) Exploring methods and techniques for extracting links from linkbases considering that content of the linkbase itself should be known in advance when associated with a specific resource for the loading of the linkbase to have any effect.
5) Exploring the possibilities of improving the approach using artificial intelligent techniques for dynamically presenting similar XLinks related to a particular topic of interest. This is a challenge of Linked Open Data as well, where a mechanism for moving forward/backwards between entities (things) should be provided especially when data sources are in the thousands, millions or billions.

The future directions mentioned above are driven irronically from XLink's flexibility and openness as it initiates wildly different implementations, which may or may not inter-operate. This paper tends to smoothens those gaps by proposing simple, straightforward approach which utilizes underlying technologies that are widely adopted by web browsers.

References

[1] C. Bizer, T. Heath, and T. Berners-Lee, *Linked Data: The Story So Far*. International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, pp. 1-22, 2009.

[2] S. DeRose, E. Maler, D. Orchard and N. Walsh (eds.) *XML Linking Language (XLink) version 1.1.* W3C Recommendation, 2010. http://www.w3.org/TR/xlink11/

[3] M. Kay (eds.) *XSL Transformations (XSLT) version 2.0*. W3C Recommendation, 2007, http://www.w3.org/TR/xslt20/

[4] B. G. Christensen, F. A. Hansen and N. O. Bouvin. *Xspect: Bridging Open Hypermedia and XLink*. In proc. of the Twelfth International World Wide Web Conference, 2003.

[5] Amaya W3C editor, 2012. http://www.w3.org/Amaya/

[6] F. Bauer and M. Kaltenbck, *Linked Open Data: The Essentials*. http://www.semantic-web.at/LOD-TheEssentials.pdf

[7] A. Haller, T. Groza. *A Novel Approach for Interacting with Linked Open Data*. Lecture Notes in Computer Science Volume 6997, 2011, pp 338-340.; 2011

[8] T. Berners-Lee, *Linked Data: Design Issues*. 2009. http://www.w3.org/DesignIssues/LinkedData.html

[9] Ph. Duffy, *Pro SharePoint with jQuery*. 2012, Book Section. Apress

[10] *The Linking Open Data cloud diagram*. 2011. [Accessed: March 2013] http://lod-cloud.net/

[11] A. Di Iorio, G. Montemari, and F. Vitali. *Beyond proxies: XLink support in the browser*, In the Proceedings of ITA 5 (2005).

[12] *Webmaster Guidelines: Best practices to help Google find, crawl, and index your site*, Google, 2013, [Accessed: June 2013] http://support.google.com/webmasters/bin/answer.py?hl=en&answer=35769#3

[13] I-Ching. Hsu, *SXRS: An XLink-based Recommender System using Semantic Web technologies*, Expert Syst. With Applications, vol. 36, no. 2, pp. 37953804, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2008.02.062., http://www.sciencedirect.com/science/article/pii/S0957417408001449. [Accessed June 17, 2013].

[14] Th. L. Maussam, and O. Etzioni, *Entity Linking at Web Scale*. The Knowledge Extraction Workshop (AKBC-WEKEX) at NAACL, Montreal, Canada. June 2012.

[15] M. Hausenblas, and M. Karnstedt, *Understanding Linked Open Data as a Web-Scale Database*. 2nd International Conference on Advances in Databases Knowledge and Data Applications (DBKDA), pp. 56-61, 2010.

[16] *INRIA: Inventors for the digital world.* [Accessed, Nov. 2013]. Available: http://www.inria.fr/en/

[17] *World Wide Web Consortium* [Accessed, Nov. 2013]. Available: http://www.w3.org

Bujar Raufi holds a PhD in Communications and Computer Technologies Computer Systems, Complexes and Networks from Technical University in Sofia, Bulgaria. Bujar's research background is in designing and implementing new approaches and directions regarding user Adaptive Intelligent Systems by fusing Data Mining and IR techniques with Semantic Web.

Xhemal Zenuni holds a PhD in Communications and Computer Technologies Computer Systems, Complexes and Networks from Technical University in Sofia, Bulgaria. Xhemal's research background is in distributed computing with special emphasis in grid systems and grid schedulling algorithms.

Artan Luma holds a PhD in Computer Science from Faculty of Contemporary Sciences and Technologies, South Esat European University, Tetovo, Macedonia. Artans's research background is in designing and implementing new approaches and directions for various cryptosystems including encryption algorithms that can be applied in different systems.